Data Augmentation through Luminance Transformation

Sang-Geun Choi, Chanil Park, Sooyoung Cho, Chae-Bong Sohn

Abstract*:*In deep learning, learning a model for data classification requires a large amount of data. Deep Learning has been applied to many areas over the years. In many areas, the amount of data is increasing for more accurate model learning. However, there are areas where data acquisition is difficult or limited. We will present a data augmentation method to solve this problem. In addition, it will lower the sensitivity to the light of the network via a data augmentation through a luminance transformation. We selected the YOLO v2 model to identify and compare the results through the proposed method. YOLO v2 is a version that improves both performance and speed over the previous YOLO model. The datasets for learning was PASCAL VOC 2011. For the comparison of the results, we will train the network by each dataset that performed data augmentation and those that were not. We will compare the performance of the two networks through three indicators. First, we will compare the change of loss according to epoch. Next, the accuracy of the network through the test set. Finally, we will see the results of the Class Activation Map (CAM)Data augmentation was performed by modifying luminance in this paper. The dataset used in the experiment was PASCAL VOC 2011 dataset, which consists of 28952 images and consists of 20 classes in total. The results obtained by training 50 epochs of network are compared. As a result of the test, the mAP of the general dataset is 0.6, and the proposed dataset is 0.7. In the test with the adjusted image of the light source, the mAP of the general dataset was 0.6, and the proposed method was 0.7. In the case of a network using a general dataset, the object is not recognized even when the image is detected through a different light source. In the CAM results, it was confirmed that the strength of catching features in the two networks was slightly different. In this paper, we confirmed that network accuracy improves when data augmentation is performed. Furthermore, the sensitivity of the network to light is also reduced. Based on these results, it is expected that the data augmentation will enable the more accurate network implementation in the area where the dataset is insufficient, and the performance when the data augmentation method is changed according to the experimental environment is expected to develop

Index Terms*: *Deep Learning, YOLO v2, Data Augmentation, Luminance Transformation, Class Activation Map (CAM)

Revised Manuscript Received on May 23, 2019.

Sang-Geun Choi, Department of Electronics and Communications Engineering, Kwangwoon University, 20 Gwangwoon-Ro, Nowon-Gu, Seoul, 01897, Korea

Chanil Park, Department of Electronics and Communications Engineering, Kwangwoon University, 20 Gwangwoon-Ro, Nowon-Gu, Seoul, 01897, Korea

Sooyoung Cho, Department of Electronics and Communications Engineering, Kwangwoon University, 20 Gwangwoon-Ro, Nowon-Gu, Seoul, 01897, Korea

Chae-Bong Sohn, Department of Electronics and Communications Engineering, Kwangwoon University, 20 Gwangwoon-Ro, Nowon-Gu, Seoul, 01897, Korea

I. INTRODUCTION

The use of deep learning technology has been spreading for automatic classification in the field of image recognition. Deep learning has a high degree of model complexity and it ensures high performance when a large amount of data is learned. Therefore, it requires large amount of data for deep learning [1]. As the number of applications of deep learning increases, there are some areas where the amount of data is sufficiently guaranteed. Conversely, there are also other areas that are not. Even if the amount of data is enough, there is a lack of processing for the elements coming from the external environment.

In this paper, we propose a method to solve the problems caused by the amount of light among external environmental factors and to increase the amount of data. Although there are many ways of data augmentation, the method of creating new data by adjusting the luminance value of the image is selected to reduce the sensitivity to light. By adjusting luminance of the image, the image becomes lighter or darker reducing the sensitivity to light. In order to verify the proposed method, we selected the YOLO v2 model and proceeded the learning process. YOLO v2 is an improved model in terms of detection performance and computing speed when compared with the previous version, YOLO [2]. The dataset used in the learning process was the PASCAL VOC 2011 dataset. This dataset contains approximately 30,000 images and has a total of 20 classes. 50% of the datasets were used as training sets and validation sets, and the remaining 50% were used as test set [3]. In this experiment, we used an image obtained by adjusting the indoor light source and the external light source in the laboratory environment for additional testing to check the sensitivity to the light source. In addition, Class Activation Map (CAM) was also used to ensure the performance of the network. By using the CAM, it is possible to determine the feature of the image when the network detects the object [4, 5].

In chapter 2, we will explain the contents of YOLO v2 used in this experiment, the method of data augmentation, and CAM. Then chapter 3 will present the results of experiment and analyze them. Finally, in chapter 4, we will describe the expected effect of this experiment.

*Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication*

II. MATERIALS AND METHODS

A. YOLO v2

The real-time object detector YOLO classifies the class by extracting the bounding box of the object from the images. In order to improve the performance of YOLO, various functions were added and removed, and the YOLO v2 was born [2].

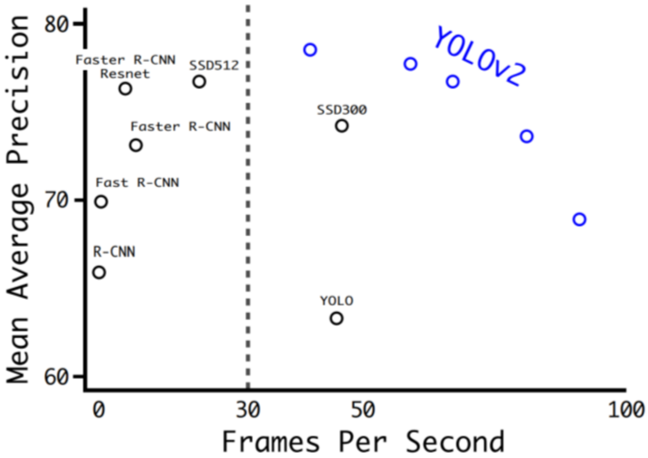

Figure 1. Accuracy and speed on VOC 2007

Fig 1 compares accuracy and speed to YOLO, YOLO v2 and other networks in VOC 2007. In VOC 2007, YOLO achieved 63.4mAP and YOLO v2 was 78.6mAP, showing a performance improvement of approximately 23% [3].
Table 1 shows the structure of Darknet-19 used in YOLO v2. In this paper, we used 20 filters in the final convolution layer (Convolution 7 in Table 1) because we were using VOC dataset.

B. Data Augmentation

Data augmentation is conducted for more accurate training. Because of the possibility of recognizing a computer as a different object even with a small rotation of the image [6]. Another reason to use is when there is insufficient dataset for network training. There are various methods of data augmentation such as image flip, crop and rotation [7, 8]. In this paper, we will perform data augmentation by adjusting the luminance value in order to reduce the sensitivity from the external light source.

C. Luminance Transformation

The images of PASCAL VOC datasets are in RGB format. Color space conversion is required to adjust the luminance of the image. Equation (1) is an expression for converting RGB format to YUV format [9, 10].

(1)

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix}$$

Adjusting the Y value in an image in YUV format allows adjustment of the luminance. After finishing the luminance adjustment, transform the YUV format image into RGB format, which is the original format, using equation (2) [9, 10].

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.164 & 0 & 1.596 \\ 1.164 & -0.391 & -0.813 \\ 1.164 & 2.018 & 0 \end{bmatrix} \cdot \left(\begin{bmatrix} Y \\ U \\ V \end{bmatrix} - \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \right) \quad (2)$$

Fig 2 is a flow chart of the entire data augmentation process. We performed data augmentation by increasing and decreasing the value of luminance by 30% and 60%.

Table 1. Darknet-19 Architecture

Type	Filters	Size/Stride	Output
Convolutional 1	32	3 X 3	224 X 224
Max pooling 1		2 X 2 / 2	112 X 112
Convolutional 2	64	3 X 3	112 X 112
Max pooling 2		2 X 2 / 2	56 X 56
Convolutional 3-1	128	3 X 3	56 X 56
Convolutional 3-2	64	1 X 1	56 X 56
Convolutional 3-3	128	3 X 3	56 X 56
Max pooling 3		2 X 2 / 2	28 X 28
Convolutional 4-1	256	3 X 3	28 X 28
Convolutional 4-2	128	1 X 1	28 X 28
Convolutional 4-3	256	3 X 3	28 X 28
Max pooling 4		2 X 2 / 2	14 X 14
Convolutional 5-1	512	3 X 3	14 X 14
Convolutional 5-2	256	1 X 1	14 X 14
Convolutional 5-3	512	3 X 3	14 X 14
Convolutional 5-4	256	1 X 1	14 X 14
Convolutional 5-5	512	3 X 3	14 X 14
Max pooling 5		2 X 2 / 2	7 X 7
Convolutional 6-1	1024	3 X 3	7 X 7
Convolutional 6-2	512	1 X 1	7 X 7
Convolutional 6-3	1024	3 X 3	7 X 7
Convolutional 6-4	512	1 X 1	7 X 7

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Convolutional 6-5	1024	3 X 3	7 X 7
Convolutional 7	20	1 X 1	7 X 7
Average pooling		Global	1000
Softmax			

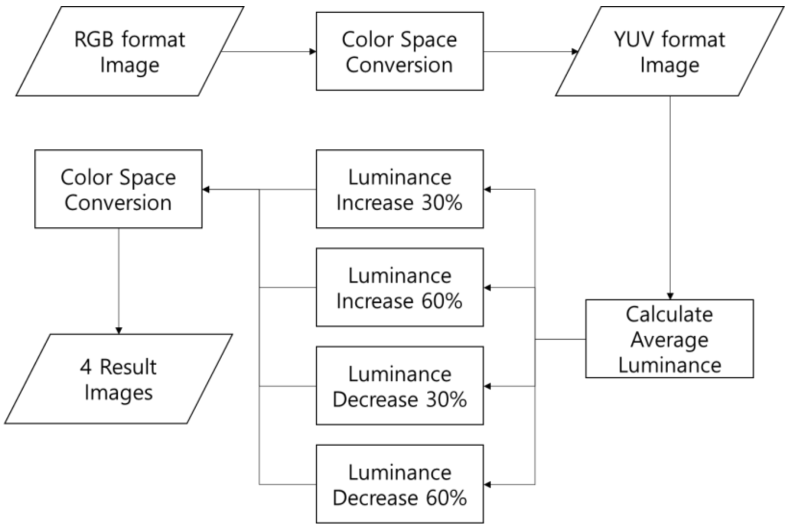

Figure 2. Data Augmentation Process

D.Class Activation Map

Class Activation Map (CAM) is a way to visualize the feature map of a deep learning. Through the CAM, it is possible to determine which feature of the image the network has assigned to the class. CAM extracts the feature map of the last convolution layer and calculates it [4, 5].

$$M_c(x,y) = \sum_k w_k^c f_k(x,y) \quad (3)$$

Equation (3) is the expression for obtaining the CAM. The elements of the equation are summarized in Table 2.

Table 2. Elements of equation 3

M	Class Activation Map
c	Detected Class
(x,y)	x, y coordinates of feature map
k	Channel
w_k^c	The k-th weight value of class c
f_k	The k-th feature map

Fig 3 is a graphical representation of (3). The value of each channel of the feature map is multiplied by the corresponding weight, and the sum of the results is used to obtain the CAM result.

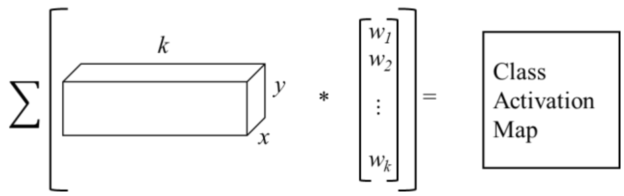

Figure 3. Class Activation Map

III. RESULTS AND DISCUSSTION

On the quantitative side of the dataset, when data augmentation was performed by increasing and decreasing the luminance value by 30% and 60%, the dataset could be increased five times that of the original dataset. Experiments were conducted to determine the differences in learning through an original dataset and the dataset proceed with data augmentation. The network used for learning was YOLO v2, and VGA used GeForce GTX 1080Ti. The training progressed over a total of 500 epochs.

In order to compare two networks, we first compared the results through the test set provided by PASCAL VOC. Table 3 shows the mAP results for both networks.

Table 3. The mAP of the trained model

Method	mAP
Original	69.7
Proposed	76.4

From the results of Table 3, we observed that the performance of the network learning the proposed dataset improved about 8% even in the test set of the general environment.

Fig 4 displays the results of the CAM. When the two networks detect the bird in the picture, we can see that the range of heat map is different.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

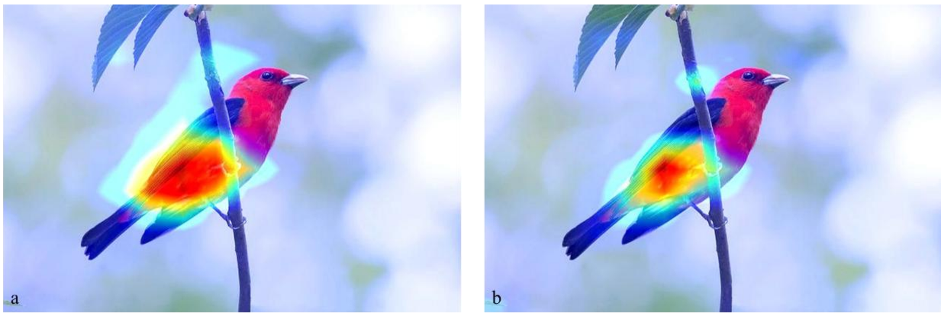

Figure 4. Class Activation Map Result: (a) Original dataset (b) Proposed dataset

When the proposed dataset is used, it can be seen that the feature density is high and the detection of the feature of the background is small.

Fig 5 shows the test results when the amount of light source is different in the laboratory environment. Among the classes provided by the VOC, we selected monitors that can be collected in the laboratory environment.

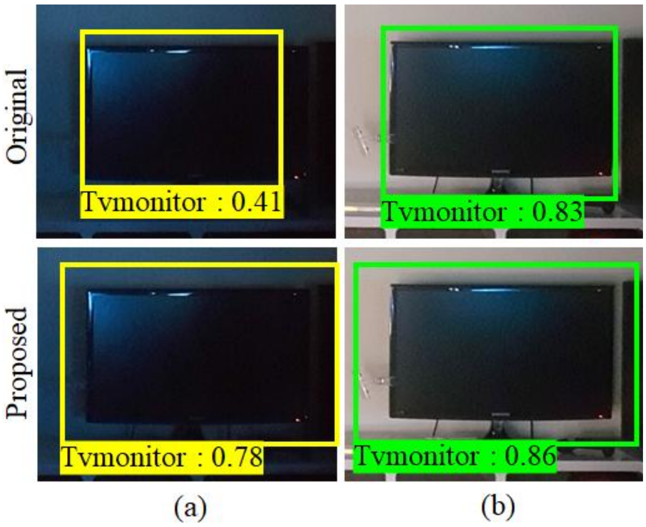

Figure 5. Detection Result: (a) Low Brightness (b) High Brightness

The results of Fig 4 show that the accuracy of the network using dataset with data augmentation in both low and high brightness is higher. Moreover, it was confirmed that the difference in accuracy at the low brightness was larger. These results demonstrate that the sensitivity of the network to light is reduced.

IV. CONCLUSION

In this paper, the data augmentation was performed by adjusting the luminance value to train the network. By increasing and decreasing the value of luminance by 30% and 60%, the amount of data increased by 5 times, and it was possible to fill less dataset. We trained original dataset and proposed dataset on each network for performance verification and compared the results. The results using VOC test set showed better accuracy in the proposed dataset. Moreover, the results of the test image with different amounts of light source in the laboratory environment were able to detect more accurately in the proposed dataset. Subsequently, the CAM results show that the proposed dataset is more robust.

In conclusion, we can confirm that adjusting the value of luminance reduces the sensitivity of the network to external light sources. Based on the results of this paper, it is expected that more accurate network training will be possible when data augmentation is performed considering the environment in which deep learning is used.

ACKNOWLEDGMENT

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2019-2016-0-00288) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation) and The present Research has been conducted by the Research Grant of Kwangwoon University in 2018.

REFERENCES

1. HE, Kaiming, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 770-778.
2. REDMON, Joseph; FARHADI, Ali. YOLO9000: better, faster, stronger. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 7263-7271.
3. EVERINGHAM, Mark, et al. The pascal visual object classes (voc) challenge. International journal of computer vision, 2010, 88.2: 303-338.
4. ZHOU, Bolei, et al. Learning deep features for discriminative localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 2921-2929.
5. KUMAR, Devinder, et al. Understanding Anatomy Classification Through Attentive Response Maps. arXiv preprint arXiv:1611.06284, 2016.
6. ZHONG, Zhun, et al. Random erasing data augmentation. arXiv preprint arXiv:1708.04896, 2017.
7. TAYLOR, Luke; NITSCHKE, Geoff. Improving deep learning using generic data augmentation. arXiv preprint arXiv:1708.06020, 2017.
8. PEREZ, Luis; WANG, Jason. The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621, 2017.
9. KITAMURA, John; THUT, Andreas; LAKSONO, Indra. Optimized color space conversion. U.S. Patent No 5,923,316, 1999.
10. YANG, Yang; YUHUA, Peng; ZHAOGUANG, Liu. A fast algorithm for YCbCr to RGB conversion. IEEE Transactions on Consumer Electronics, 2007, 53.4: 1490-1493.

AUTHORS PROFILE

Sang-Geun Choireceived B.S. degree in Electronics and Communications Engineering from Kwangwoon University in 2015. He is currently a Ph.D. candidate in department of Electronics and Communications Engineering, Kwangwoon University, Seoul, Korea.

Chanil Parkreceived B.S degree in Electronics and Communications Engineering from Kwangwoon University in 2019. He is currently a M.S.student in department of Electronics and Communications Engineering, Kwangwoon University, Seoul, Korea.

Sooyoung Choreceived B.S degrees in Daegu Haany University in 2013. He is currently a Ph.D. candidate in department of Electronics and Communications Engineering, Kwangwoon University, Seoul, Korea

Chae-Bong Sohnreceived the B.S., M.S., and Ph.D. degree in Electronic Engineering from Kwangwoon University in 1993, 1995, and 2006, respectively. He is currently an associate professor in department of Electronics and Communications Engineering, Kwangwoon University. His research interests include image and video processing, NLP systems, and A.I. systems

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication